

# 日本語第二言語学習者を対象とした 心理言語学的語彙評定データセット

鄭弯弯<sup>1</sup>

<sup>1</sup> 名古屋大学大学院人文学研究科附属人文知共創センター  
zheng@nagoya-u.jp

## 概要

語彙習得は、単語の表記や音声といった表層的な形式の学習にとどまらず、語の意味処理に関する認知的特性や、学習者の経験・心理状態が関与する複合的な過程である。にもかかわらず、日本語を含む多くの言語における語彙難易度研究では、出現頻度や文字数などの表層的指標が主に用いられてきた。一方で、語が直感的に「理解しやすい」「理解しにくい」と感じられる背景にある認知的・情動的要因については、十分な検討がされていない。この課題に対応するため、本研究では、日本語第二言語学習者を対象とした大規模調査を実施し、親密度、心像性、抽象度、感情価、覚醒度、理解しやすさという六つの心理言語学的次元に基づく語彙評定データセットを構築した。本データセットは、語彙難易度をより精密な推定を可能にするだけでなく、視覚的・感情的・認知的・文脈的要因の相互作用を通じた語彙処理の形成過程を体系的に検討するための研究資源を提供する。

## 1 はじめに

実証研究は、単語を正確かつ迅速に認識できることが、文章内容を理解するための基礎条件であることを一貫して示している [1, 2, 3, 4]。また、語彙被覆率は、理解可能な情報の範囲を左右する要因であるだけでなく、一般的な認知能力を反映する比較的安定した指標としても報告されている [5]。加えて、就学前児を対象とした縦断的・発達の研究からは、実行機能（注意制御や抑制、切り替え能力など）と語彙量が相互に影響し合いながら発達することが示されている [6]。これらの知見を踏まえると、語彙知識は単に言語的記号を蓄積するものではなく、効率的な言語処理を可能にすると同時に、認知制御の発達を支える重要な基盤であると位置づけられる

[7, 8]。

一方で、語の難易度に関する情報は、限られた学習時間の中で語彙を効率的に習得するための重要な指標となる。しかし、語の難易度は単純に定義できるものではなく、語の長さ [9] や形態素数といった形態的特徴、多義性 [10, 11] や特定分野への偏り [12] といった意味的特徴、出現頻度 [13] や親密度 [14, 15] などの学習者の接触経験、語がどの程度具体的な心的イメージを喚起するかを示す心像性・具体性 [16, 17]、さらに情動 [18, 19, 20]、認知処理の容易さ [21, 22]、記憶保持のしやすさ [23, 24, 25] といった心理的要因が複合的に関与している。

多種多様な影響要因の中に、統計的指標は、計算や解釈が比較的容易であり、大規模分析に適しているという利点を持つが、その値は使用するコーパスや辞書型データベースの性質に強く依存する。一方で、親密度や心像性といった心理的・認知的側面に基づく語彙特性は、学習者や使用者が語をどのように経験し、処理しているかを直接的に反映する指標である。そのため、これらの特性は、頻度に基づく統計量だけでは捉えきれない語の難易度の側面を補完し、語彙処理や語彙学習をより包括的に理解するための重要な情報を提供すると考えられる。

これらの課題を踏まえ、本研究では、日本語を第二言語として学習する学習者を対象とした新たな語彙データセットを構築し、語の難易度をより精密に推定することを目的とした。本データセットは、親密度（語へのなじみの程度）、心像性（心的イメージの喚起しやすさ）、抽象度（概念の物理的・具体性の程度）、感情価（肯定的－否定的感情）、覚醒度（落ち着く－刺激的）、理解しやすさ（表層的・認知的な影響を含む統合的な評価）という六つの心理言語学的特性を統合的に収録している。これにより、単純な頻度指標では捉えにくい、語彙処理における認知的・情動的な複雑性を多角的に捉えることが可

能となる。

本稿の構成は、以下のとおりである。第2節では、調査票の設計およびデータ収集に用いた具体的な手続きを説明する。続く第3節では、得られたデータの構造的特徴を、定量的分析および可視化の結果を通して提示する。最後に、第4節において、本データをまとめる。

## 2 調査票の設計とデータ収集

### 2.1 調査語の選定

前節で述べたように、語の難易度は、言語的特性、認知的処理負荷、心理的要因の相互作用によって形成される本質的に多次元的な概念である。このような複合的要因が語の知覚や処理に影響を及ぼすことを踏まえると、正確かつ信頼性の高い語難易度情報を整備することは不可欠である。こうした情報は、妥当性の高いデータセット構築の基盤となるだけでなく、それに基づく後続分析の解釈可能性と信頼性を左右する。そこで本研究では、語の難易度評価の対象語として、『日本語教育語彙表』に収録された語彙を採用した [26]。

同辞典には 17,920 語の見出し語が含まれており、その語彙は二つの異なる性質を持つ語彙資源に基づいて構築されている。一つは、現代日本語の書き言葉全体の特徴を反映することを目的とした現代日本語書き言葉均衡コーパス (BCCWJ) に基づく語彙調査であり、もう一つは、初級から上級までの学習段階を網羅する、市販日本語教材 100 冊分の電子データから成る日本語教科書コーパスである (未公開資料)。このように、対象語が均衡の取れた全国規模のコーパスと、実際の教育現場で用いられている教材の双方から体系的に抽出されている点は、本辞典を対象語選定の基盤とする大きな利点である。その結果、得られた語彙リストは、新聞・書籍・教材など多様なジャンルやレジスターを含みつつ、第二言語学習者にとって実用性の高い語彙を広くカバーしている。

本辞典の各見出し語には、表記の揺れを統一した標準表記、綴り情報、語の難易度、品詞 (POS)、および語種 (和語・漢語・外来語・混種語) が体系的に注記されている。語の難易度は、日本語教育において十分な指導経験を有する教師による主観的判断に基づき、初級・中級・上級の三水準をそれぞれ上下に細分化した計 6 段階で設定されている。各難易

度レベルに属する代表的な語の例を表 1 に示す。

表 1 対象語の例 (出所:『日本語教育語彙表』)

標準表記	読み	難易度レベル	POS	語種
秋祭り	アキマツリ	中級後半	名詞	和語
愛好	アイコウ	上級前半	名詞	漢語
アイス	アイス	初級前半	名詞	外来語
空き缶	アキカン	中級前半	名詞	混種語

### 2.2 調査票の設計

各語に対する調査項目は、図 1 に示す統一形式で提示された。日本語では、同一の表記を持ちながら異なる読みや意味を持つ同形異音語が存在し、こうした違いが主観的な難易度評価に影響を与える可能性がある。そのため、本研究では、刺激語を必ず「語+読み」の形式で提示し、評価対象となる語の解釈が一意に定まるよう配慮した。参加者は、提示された各語について、認知的側面と情動的側面の両方を含む印象を、以下の二つの観点から評価した。

**親密度** 本項目では、対象語に対する使用経験に基づく親密度を測定した。具体的には、「書く」「読む」「話す」「聞く」という四つの使用場面における親密度と、語全体に対する総合的印象について評価を求めた。これらの評価は、「まったくそう思わない」から「非常にそう思う」までの 7 件法リッカート尺度で収集された。

**意味およびイメージ** 本項目では、対象語の意味的および心理的特性に関する主観的評価を収集した。参加者は、ネガティブさ (感情価: 快-不快)、落ち着き度 (覚醒度: 落ち着く-刺激的)、イメージの想起しやすさ (心像性)、抽象度 (概念的な難しさ)、および理解しやすさ (学習時に感じる難しさ) という五つの側面について評価を行った。これらの回答も同様に 7 件法リッカート尺度で記録された。

本調査では、各参加者に対して 7 語から構成される語彙セットを 2 セット提示し、合計 14 語について評価を求めた。対象語は、採用した辞典の語彙リストから無作為に抽出され、6 段階に分類された語難易度のすべてが均等に反映されるよう設計された。本課題は正誤判断を伴わない主観評価課題であるため、参加者には分析的判断ではなく、語に対する直感的な印象に基づいて回答するよう求めた。

回答の信頼性を検証する目的で、2 語を重複語として両セットに含め、同一語を 2 回評価させた。これらの重複語は、記憶効果や提示順序効果を最小限

「秋祭り(アキマツリ)」という語について、以下の観点から評価してください。

このフォームを送信する際、お客様が、ご自身のお名前やメールアドレスなどの詳細情報を入力しない限り、その情報が自動的に取得されることはありません。

1. 親しみやすさについて [Q]

	全く思わない	思わない	少し思わない	どちらもない	少し思う	思う	非常に思う
書く時	<input type="radio"/>						
読む時	<input type="radio"/>						
話す時	<input type="radio"/>						
聞く時	<input type="radio"/>						
全体	<input type="radio"/>						

2. 意味・イメージについて [Q]

	全く思わない	思わない	少し思わない	どちらもない	少し思う	思う	非常に思う
ネガティブさ	<input type="radio"/>						
落ち着き	<input type="radio"/>						
イメージしやすい(空からの場面やものが思い浮かぶ)	<input type="radio"/>						
抽象的(空目に見えない考えや概念)	<input type="radio"/>						
理解しやすい	<input type="radio"/>						

図 1 調査表

に抑えるため、一定の間隔を空けて配置された。語の提示順は、A セット (5 語) → 重複語 C (2 語) → B セット (5 語) → 重複語 C (2 語) という固定形式に従った。

加えて、質問紙の最後では、性別、年齢、国籍、ならびに日本語学習レベルに関する属性情報を収集した。学習レベルについては、初級から上級までの 5 段階を設定し、各段階に具体的な説明文を付すことで、参加者が自身の運用能力に最も近い選択肢を選びやすいよう配慮した。

## 2.3 調査の実施とスクリーニング

**調査の実施** 本調査の質問紙は Qualtrics を用いて設計され、URL または QR コードを介してオンライン形式で実施された。研究の実施にあたっては、大学倫理審査委員会の承認を受けており (承認番号: NUHM-25-17)、すべての参加者から調査開始前にオンライン上でインフォームド・コンセントを取得した。

調査は 2025 年 8 月～12 月 31 日に実施された。回答に要する時間は約 10～18 分であった。調査の冒頭では、研究の目的および内容を説明した同意文書が提示され、同意が確認された後にのみ回答データが記録された。謝礼として、有効な回答を提出した参加者には 300 円相当の電子ギフトカードが提供され、中国在住の参加者には 15 人民元相当の謝礼が支給された。十分なサンプル数を確保するため、同一参加者による複数回の参加を認め、回答回数は

最大 10 回までとした。

**スクリーニング** 回答データの信頼性を確保するため、まず欠損値を含むエントリを除外した。続いて、各参加者が重複して評価した 2 語について、1 回目と 2 回目の評価結果を比較した。リッカート尺度による主観評価 (「まったくそう思わない」～「非常にそう思う」) は、分析のために 0～6 の数値尺度に変換された。

回答の一貫性は、平均絶対誤差 MAE と完全一致率 Agree という二つの補完的な指標を用いて評価した。Agree は、同一の評価が得られた割合を示す直感的な指標であるが、評価のずれの大きさや方向を考慮しないという制約がある。たとえば、±1 程度の小さな差と ±3 以上の大きな差は同等に「不一致」として扱われる。そのため、本研究では、評価差の大きさを連続量として捉える MAE を併用し、回答の安定性をより精緻に把握した。

具体的には、重複語  $i = 1, 2$  に対して、各評価項目  $j = 1, \dots, 10$  の数値評価を、1 回目の評価を  $a_{ij}$ 、2 回目の評価を  $b_{ij}$  と定義した。いずれかに欠損値を含む項目は、該当する計算から除外した。

### Mean Absolute Error (MAE)

$$MAE_i = \frac{1}{10} \sum_{j=1}^{10} |a_{ij} - b_{ij}| \quad (1)$$

値の範囲は 0～6 であり、0 に近い値ほど、2 回の評価間の一貫性が高いことを示す。

### Exact Agreement Rate

$$Agree_i = \frac{1}{10} \sum_{j=1}^{10} 1(a_{ij} = b_{ij}) \quad (2)$$

値の範囲は 0～1 であり、1 はすべての項目において完全に一致していることを示す。

重複して提示された 2 語に対する評価の安定性を総合的に把握するため、MAE と一致率の平均値を算出した。人間の判断は本質的に確率的であり、同一語に対する評価が常に完全に一致することは期待できない。このような人間判断の変動性を前提とした上で、本研究では、評価のずれが 1 段階以内に収まっているか (MAE)、および完全一致がどの程度の割合で生じているか (Agree) を基準として、回答の信頼性を以下の三段階に分類した。

- **高信頼性:**  $MAE \leq 1.0$  (平均的な評価差が 1 尺度以内) and  $Agree \geq 0.70$  (7 割以上で完全一致)。
- **要再検討 (グレーゾーン):**  $1.0 < MAE \leq 2.0$  or  $0.20 \leq Agree < 0.70$ 。

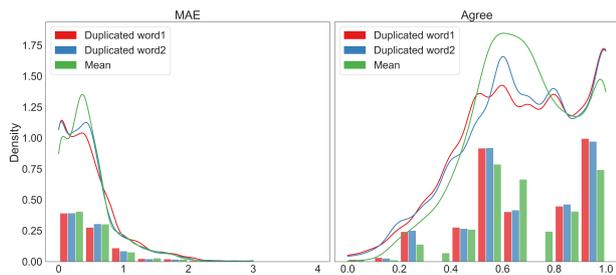


図2 重複語に対する MAE および一致率と、それらの平均値の分布

- 低信頼性: 上記条件を満たさない場合

### 3 データ構造

最終的に、17 か国にわたる 3615 名の参加者から有効回答が収集された。参加者は、中国および台湾地域、ベトナム、韓国、アメリカ合衆国、フィリピン、イギリス、ミャンマー、タイ、トルコ、フランス、インドネシア、マレーシア、モンゴル、オーストラリア、ニカラグア、バングラデシュ、インドの日本語学習者で構成されていた（回答者数の多い順）。そのうち、中国および台湾地域出身の回答者が最も多く、全体の 93.88% を占めていた。信頼性基準に基づく分類の結果、1840 件 (50.90%) が高信頼性、1775 件 (49.10%) が要再検討として扱われた。

図2は、重複語に対する2回の評価から算出された MAE と一致率の分布、およびそれらの平均値を示している。ヒストグラムは各指標の頻度分布を示し、重ねて描画されたカーネル密度推定 (KDE) 曲線は分布の形状や最頻領域を可視化している。MAE は左に歪んだ分布を示し、多くのケースが 1.0 未満に集中していたことから、同一語に対する2回の評価差は概して小さいことが分かる。一方、一致率は 0.6 付近と 1.0 付近にピークを持つ二峰性分布を示した。これは、多くの参加者において回答の一貫性が高い水準で維持されていたことを示唆している。

one-class SVM により外れ値を除外した後の平均回答時間は、675.82 秒 (正常回答数 = 3427) であった。事前に実施したパイロット調査で想定された適正範囲 (600~1080 秒) と比較すると、この値は該当範囲と整合していた。これらの結果から、本調査における回答時間は全体として事前想定と一致しており、回答時間に基づく品質確認の観点からも、参加者の大多数が妥当な時間をかけて回答していたことを示している。

参加者の年齢は 12 歳から 63 歳に分布しており、平均年齢は 25.23 歳、標準偏差は 4.03 であった。性別構成は、女性 2074 名 (57.37%)、男性 1500 名 (41.49%)、その他 41 名 (1.13%) であった。自己評価による日本語熟達度の分布を表2に示す。全体の 88% が中級から上級に集中していることから、本調査の参加者の多くは、基礎段階を超え、一定水準以上の日本語運用能力をすでに獲得している学習者であった。

表2 参加者の日本語熟達度分布

日本語熟達度	人数	割合 (%)
中級	2107	58.28
中級の上	798	22.07
上級	277	7.66
初級	254	7.03
初級の上	179	4.95

### 4 おわりに

日本語は、習得が難しく複雑な言語の一つとしてしばしば位置づけられているが、語の難易度に影響を与える要因や、それぞれの寄与の大きさについては、いまだ十分に解明されていない。

これまでの研究は主としてコーパスに基づく統計的指標に依拠してきたため、語彙処理の背後にある認知的複雑性を直接的に捉えるデータには大きな不足があった。本研究では、この課題を踏まえ、親密度、心像性、抽象度、感情価、覚醒度、理解しやすさという六つの心理言語学的次元に焦点を当て、実証的調査に基づく新たなデータセットを構築した。予備的な定量分析の結果から、本データセットは既存の理論的枠組みや先行研究の知見と整合的であり、その妥当性が示されている。

語の難易度は、表層的な語彙特性に限定されるものではなく、文脈的条件、社会文化的背景、認知的処理コスト、さらには情動的側面によっても形成される。また、これらの要因は相互に影響し合いながら語彙処理に関与しており、その全体像は依然として十分に解明されていない。本研究で構築されたデータセットは、語の難易度推定に資するだけでなく、これら多様な要因間の関係性や相互作用を体系的に検討するための基盤として、今後の理論的・実証的研究に貢献することが期待される。

## 謝辞

本研究は JSPS 科研費 25K16328、JPJS00122674991 の助成を受けたものである。また、本データセットの収集にあたり、名古屋大学 大平英樹先生、同志社大学 李長波先生、大連外国語大学 孫昊先生、暨南大学 李広微先生をはじめ、お名前の存じ上げない日本語学校のご担当者の皆様に多大なご協力を賜った。ここに記して、深く感謝の意を表する。

## 参考文献

- [1] Fiona J. Duff, Gurpreet Reen, Kim Plunkett. Do infant vocabulary skills predict school-age language and literacy outcomes? *Journal of Child Psychology and Psychiatry and Allied Disciplines*pp. 848–856, 8 2015.
- [2] Callula Killingly, Linda J. Graham, Haley Tancredi. Reciprocal relationships among reading and vocabulary over time: a longitudinal study from grade 1 to 5. *Reading and Writing*pp. 605–625, 3 2025.
- [3] Walter Kintsch. *Comprehension: A Paradigm for Cognition*. Cambridge University Press, Cambridge, UK, 1998.
- [4] C.A. Perfetti. Comprehending written language: A blueprint of the reader, pp. 167–208. *The neurocognition of language*. Oxford University Press, New York, NY, 1999.
- [5] Madeleine Bruce. Vocabulary and executive functioning: A scoping review of the unidirectional and bidirectional associations across early childhood. *Human Development*pp. 167–187, 8 2022.
- [6] Sara A. Schmitt, David J. Purpura. Predictive links among vocabulary, mathematical language, and executive functioning in preschoolers. *Journal of Experimental Child Psychology*pp. 55–68, 4 2019.
- [7] J. D. Bransford, A. L. Brown. *How people learn: Brain, mind, experience, and school*. National Academy Press, Washington D.C., U.S., 2000.
- [8] Regina Jucks. What makes a word difficult? insights into the mental representation of technical terms. *Metacognition and Learning*pp. 91–111, 8 2012.
- [9] Gina N. Cervetti, Elfrieda H. Hiebert, P. David Pearson. Factors that influence the difficulty of science words. *Journal of Literacy Research*pp. 153–185, 12 2015.
- [10] Joseph P. Vitta, Christopher Nicklin. Academic word difficulty and multidimensional lexical sophistication: An english-for-academic-purposes-focused conceptual replication of hashimoto and egbert (2019). *Modern Language Journal*pp. 373–397, 3 2023.
- [11] Derek N. Canning, Stuart McLean. Relative complexity in a model of word difficulty: The role of loanwords in vocabulary size tests. *Studies in Second Language Learning and Teaching*pp. 631–659, 12 2024.
- [12] Erina Yamakawa. Estimating word difficulty using wikipedia (in japanese), 2024.
- [13] Brett J. Hashimoto. More than frequency? exploring predictors of word difficulty for second language learners. *Language Learning*pp. 839–872, 12 2019.
- [14] GONDY Leroy. The effect of word familiarity on actual and perceived text difficulty. *Journal of the American Medical Informatics Association*, 2014.
- [15] Wanwan Zheng. Estimating word difficulty using stratified word familiarity. *Cogent Arts and Humanities*, 2024.
- [16] M. Marschark. *Imagery and Verbal Memory*, pp. 133–182. *Imagery and cognition*. Springer, New York, US, 1991.
- [17] Alper Kumcu. Less imageable words lead to more looks to blank locations during memory retrieval. *Psychological Research*pp. 667–684, 4 2020.
- [18] J.M. Dewaele. From obscure echo to language of the heart: Multilinguals’ language choices for (emotional) inner speech. *Journal of Pragmatics*pp. 1–17, 2015.
- [19] Mara Mather. Arousal-biased competition in perception and memory. *Perspectives on psychological science : a journal of the Association for Psychological Science*pp. 114–133, 2011.
- [20] Tae-Ho Lee, Michiko Sakaki, Ruth Cheng, Ricardo Velasco. Emotional arousal amplifies the effects of biased competition in the brain. *Social Cognitive and Affective Neuroscience*, pp. 2067–2077, 03 2014.
- [21] Yu Kanazawa. Replication and extension of an empirical study about foreign language formulaic familiarity database: Reliability and Emotionality, pp. 149–164. *Bulletin suisse de linguistique appliquée*. 2021.
- [22] Yu Kanazawa. Lexical and contextual emotional valence in foreign language vocabulary retention. *The Mental Lexicon*pp. 339–365, 12 2023.
- [23] Deborah Talmi, Ulrich Schimmack, Theone Paterson. The role of attention and relatedness in emotionally enhanced memory. *Emotion (Washington, D.C.)*pp. 89–102, 02 2007.
- [24] John Truscott. Consciousness in sla: A modular perspective. *Second Language Research*, pp. 413–434, 2015.
- [25] L. F. Barrett. *How emotions are made: The secret life of the brain*. Houghton Mifflin Harcourt, Boston, US, 2017.
- [26] Y. Sunakawa, J. Lee. The construction of a database to support the compilation of japanese learners dictionaries. *Acta Linguistica Asiatica*, pp. 97–115, 2012.